



Sullivan, B., Wilson, C. E., & Saldaña, D. (2019). Development of a gaze contingent method for auditory threshold evaluation in non-verbal ASD children. *Research in Autism Spectrum Disorders*, 62, 85-98. <https://doi.org/10.1016/j.rasd.2019.02.006>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.rasd.2019.02.006](https://doi.org/10.1016/j.rasd.2019.02.006)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.rasd.2019.02.006> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

1
2
3
4
5
6
7
8
9
10
11
12 **Development of a gaze contingent method for auditory threshold evaluation in non-**
13 **verbal ASD children**

14
15 Brian Sullivan^{1,3*c}, C. Ellie Wilson^{2*}, David Saldaña²
16

- 17 1. School of Psychological Science, University of Bristol, UK
18 2. Individual Differences, Language and Cognition Lab, Department of Developmental and
19 Educational Psychology, University of Seville, Spain
20 3. Tobii AB, Danderyd, Sweden
21

22 *Authors contributed equally to the research and manuscript
23

24 °Corresponding Author:
25 School of Psychological Science
26 12A Priory Road, Bristol, UK, BS8 1TU
27 brian.sullivan@bristol.ac.uk
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 **Keywords:** children, audition, non-verbal, eye tracking, gaze contingent, perceptual threshold
51

Abstract

Background: Minimally verbal children with autistic spectrum disorder (ASD) make up an estimated third of the ASD population (Downs, Schmidt, & Stephens, 2005), but have been understudied due to difficulties in running experiments with such participants. We sought to develop an instrument to evaluate auditory perception, with the goal of testing both typically developing (TD) and ASD children, including minimally verbal ASD. Audio difference thresholds are typically measured by an audiologist using visual reinforcement audiometry (VRA) techniques, but this requires a trained clinician. Alternatively, mismatch negativity (MMN) via an electroencephalogram can provide an objective threshold measure and the participant can passively attend to stimuli. However, EEG equipment is expensive, and the procedure can be uncomfortable and difficult with anxious or touch sensitive participants.

Method: We developed a testing software for estimating auditory thresholds in children using a gaze contingent 'game.' Our open source software uses an eye tracker, Matlab and child-oriented stimuli to automate aspects of VRA. Initial results suggest that audio thresholds can be obtained using our affordable non-invasive system, operated with minimal training, but refinement is necessary.

Results: Our method can obtain thresholds for most typical children, but data collection in young ASD children proved more challenging, yielding poor results, and will require further development to make the game more accessible. While promising, these results need to be corroborated with an alternate measure of difference threshold.

Conclusion: We document our efforts to design an effective interactive game to assess auditory perception using gaze-contingent eye-tracking methods; and provide case level insights on the testing individual participants and the heterogeneous ability and performance levels within ASD. We discuss the challenges experienced in testing and eye tracking both typical and ASD children to inform clinical and research groups to advance this promising line of research.

Introduction

Background and Motivation

Linguistic ability within autism spectrum disorder (ASD) can range from fluent speech with average (or above) verbal intelligence, to 'minimally verbal', defined here as an individual with fewer than five productive words by the time they reach school age (Tager-Flusberg & Kasari, 2013). It is estimated that 30% of ASD children remain minimally verbal (Tager-Flusberg & Kasari) and it is this severely impaired end of the spectrum that is arguably most in need of attention. Yet the majority of research in ASD focuses on children with age-appropriate or moderately-impaired language, and it is questionable whether insights gained from these studies may be extrapolated to those who are minimally verbal. A primary reason for the lack of research into the 'neglected end of the spectrum' is that there are substantial challenges involved in completing tests with these individuals. As a result, a priority for ASD research today is to develop evaluation techniques using available technology that can effectively assess cognitive and behavioral traits of individuals across the autistic spectrum.

Assessing receptive and expressive language is important but targeting fundamental traits that might underlie linguistic difficulties is also critical. Atypical sensory behaviors – including hyper-sensitivities, hypo-sensitivities, or unusual sensory seeking behaviors – are a common characteristic of ASD with prevalence estimates between 75% (Klintwall et al., 2011) and 90% (Crane et al., 2009). Perceptual experiences are acquired early and form the basis of our interaction with the environment, therefore atypical perception of auditory information may have knock-on effects for the development of more advanced functional domains, such as speech and language (Stevenson et al., 2014; Watson et al., 2012). Rhythm perception, for example, has been linked to skill level in speech comprehension (Drullman et al., 1994; Elliott & Theunissen, 2009; Bertoni et al., 2011; Slater et al. 2015), grammar (e.g. Gordon et al. 2015) and reading (e.g. Huss et al. 2011; Woodruff Carr et al. 2014). Rhythm perception has also been a target for therapeutic intervention in speech and language disorders such as ASD, stuttering, aphasia and Parkinson's disease (Fujii & Wan, 2014). Pitch perception, on the other hand, has repeatedly been reported as atypical (often enhanced) in at least a subgroup of ASD individuals (Bonnell et al. 2003, 2010; Jones et al. 2009; Samson et al. 2006; Stanutz et al. 2014), and has been associated with language learning in typically developing infants (Mueller et al. 2012). Furthermore, within the study of ASD auditory perception, a distinction between the perception of speech and non-speech stimuli has been highlighted as it is hypothesized that ASD children with severe impairments in social communication may have a specific impairment processing complex speech sounds. Studies demonstrating atypical responses to speech sounds in particular have been documented using EEG (Kuhl et al. 2005; Yau et al. 2015; Schwartz et al., 2018) and behavioral (Heaton et al. 2008) methods.

However, of the studies mentioned here, only one (Yau et al. 2015) has assessed auditory perception in minimally verbal subjects. Moreover, studies invariably assess features of auditory perception – e.g. pitch and rhythm – in isolation. Devising tasks that can examine multiple components of auditory perception in a comparable manner is important if we are to understand which features may be most relevant to language development. For these reasons, we aimed to develop an objective measure that could examine different features of auditory perception, and that could be used for all children including those that are minimally verbal.

Audiometry and Hearing Thresholds

1
2 Within psychophysics, perception of sensory signals can be quantified by absolute detection
3 thresholds, i.e. what is the softest sound one can hear before only perceiving silence, and
4 auditory difference thresholds, i.e. what is the smallest difference one can perceive between two
5 tones before they are perceived as identical. In the current study, we focus solely on auditory
6 difference thresholds.

7
8 In hearing evaluation, there are several common ways to evaluate auditory difference
9 thresholds. Standard hearing tests require the participant to communicate with an experimenter
10 or audiologist, with the patient verbally instructed to signal when they perceive a tone or a
11 difference between tones. In nonverbal or minimally verbal populations, such as young children,
12 there are two primary methodologies to evaluate thresholds without requiring the child to
13 understand instructions or actively communicate with the researcher; one uses overt behavioral
14 measures and the other electro-encephalography (EEG) signal recording.

15 *Behavioral Methods*

16
17
18 Audiologists have developed a set of non-invasive observational techniques to present stimuli
19 and evaluate orientation and physical behavior in response to the stimulus (Sabo 1999).

20
21 In Behavioral Observation Audiometry (BOA) sounds are presented and the audiologist must
22 observe the infants (new born to ~5 months) behavior (eye and head orientation and/or facial
23 expressions) when exposed to sound stimuli.

24
25 Visual Reinforcement Audiometry (VRA) and Conditioned Orientation Reflex (COR) are
26 techniques that both rely on the child (typically 5 months – 2 years) learning a stimulus–
27 response pairing and responding with an orientation response (eye, head and/or body turn).
28 Initially, the audiologist will present an anchor sound and reward the child when an orientation
29 response is made (rewards may be, for example, shaking a toy or moving a puppet). In VRA an
30 absolute threshold may be evaluated once the child has learned to orient to a sound. The
31 audiologist then may gradually lower the volume of the sound over a series of presentations
32 until the orientation response is no longer elicited indicating the child no longer perceives the
33 sound. COR expands the same premise to have two or more sound sources, useful for auditory
34 difference thresholds. For instance, the child may be reinforced to orient to the left for one
35 sound and to the right for another. If the child cannot perceive a difference between the two
36 stimuli their pattern of choice behavior should be 50/50.

37
38 Conditioned Play Audiometry is similar to the standard audiology test but adds extra incentives
39 to keep children (2-3 years) engaged. For instance, instructing the child to use a toy in a certain
40 way if they hear a particular sound, such holding a toy block and dropping it as soon as they
41 hear a sound.

42 *EEG Methods*

43
44
45 Alternatively, auditory difference thresholds can be determined via auditory evoked potentials
46 (AEP), electrical activity from populations of synchronously firing neurons (Luck 2005), recorded
47 via EEG. This is an extremely useful method as not only can it be used to evaluate thresholds in
48 nonverbal participants, it works in participants that are not actively attending and can be
49 conducted in an automated fashion not requiring an audiologist to observe the participant
50 behavior.

1 In such studies, the timing of EEG responses can be accurately recorded and analyzed
2 according to distinct portions that match to different stages of processing (Musiek & Baran
3 2007). Early brainstem responses (0-15ms) can be used to determine absolute thresholds
4 (Paulraj et al., 2015). If a participant repeatedly hears a stimulus and then is presented with a
5 new dissimilar stimulus, a mismatch negativity (MMN) response arises late in the record (200-
6 400ms) and can be used to estimate auditory difference thresholds (Näätänen et al 2007).
7 However, MMN is problematic for clinical assessment of individual participants and typically
8 requires across subject averaging. Schall (2016) has argued that MMN can be clinically useful
9 for measuring sound processing impairments but not a final diagnostic and will require
10 substantial investment to target specific patient groups and gather normative data sets.
11 Furthermore, whilst EEG techniques are non-invasive and don't require overt behavior, the
12 testing environment can be intense, and the participant is required to endure a set-up procedure
13 involving putting a cap on their head with multiple electrodes that need to be secured. This can
14 be extremely difficult in a subject who may be touch sensitive, anxious, hyperactive and/or
15 noncompliant.

16 17 18 ***Testing Hearing in the ASD Population*** 19

20 The sensory atypicalities that are common in ASD also contribute to the difficulties in conducting
21 tests in this population. In the case of hearing testing, sensory atypicalities could mean that an
22 ASD participant dislikes wearing items like headphones, or they find the volume of sound
23 presentation overwhelming. Rosenhan et al. (1999) outlined several additional common
24 characteristics and challenges in testing the ASD population including: anxiety, hyperactivity
25 and poor attention, cognitive and language comprehension impairments, difficulty in new
26 environments, increased false positive and false negative responses, and rapid or slowed
27 habituation. Although any individual participant may only exhibit a portion of these
28 characteristics they all contribute to difficulty in collecting reliable experimental data. Similarly,
29 Plesa Skwerer et al (2016) measured receptive language ability in ASD children and found a
30 large amount of variability arguing for individually tailored assessment and interventions. Downs,
31 Schmidt, and Stephens (2005) did not recommend standard auditory exams (i.e. the patient
32 explicitly signals when a tone is heard) with the ASD population and instead advised that patient
33 history (communicated by the parent) and behavioral observation audiometry could be more
34 successful. Kasari et al (2013) recommend a comprehensive inventory of behavior in minimally
35 verbal ASD children covering both verbal and nonverbal behavior. Individualized methods could
36 certainly be useful for gaining insight into one child's ability and potential difficulty. However, this
37 approach is limited in that conclusions regarding auditory difference thresholds and the potential
38 relationship with additional cognitive and linguistic abilities in a wider population is not possible
39 when different methods are used for every participant.

40 41 ***Motivation for Current Approach*** 42

43 While assessment of thresholds via behavioral observation or auditory evoked potentials are
44 well established, they have some requirements that may be difficult to meet in a non-specialist
45 lab. Without access to an audiology clinic or auditory testing facilities many of the behavioral
46 testing options are not feasible. Similarly, without an EEG system and a trained technician to
47 collect and interpret the data, it is not clear how to best assess audition. As a result, reports of
48 EEG studies in minimally verbal ASD participants are extremely sparse (Yau et al., 2015). In our
49 case, we decided to try to convert some aspects of visual reinforcement audiometry into a
50 'game' that uses eye movements as a means of interaction. Eye tracking hardware is becoming
51 more affordable and easier to use, and there are several additional advantages that are

1 particularly beneficial when testing a minimally verbal ASD population: the set-up time is
2 minimal, physical elements like attaching electrodes are not required, testing can begin as soon
3 as calibration is achieved, and participants are permitted to make some movements during
4 testing without adversely affecting data. In children who may have impaired motor skills, the
5 eyes can be used effectively as means to select onscreen imagery, and once calibrated an eye
6 tracker can be used as an intuitive real-time interaction device.

7
8 Mixed success has been reported using gaze contingent experiments in infant participants.
9 Several researchers have demonstrated that infant eye movements can be successfully
10 recorded and infants can learn how to interact within a gaze contingent experiment (Wass,
11 Porayska-Pomsta & Johnson, 2011; Wang et al., 2012; Miyazaki et al. 2014). These studies
12 trained participants to control their visual attention by reinforcing 'correct' gaze movements with
13 visual and auditory rewards. They demonstrated that infants learn to anticipate rewards and
14 moderate their eye-movements to control their environment. The only study that we are aware
15 of that attempted to conduct an auditory test using a gaze contingent eye-tracking set-up to
16 record 'responses' is reported only as proceedings from a conference (Schwarz et al. 2014).
17 These authors attempted to train 12 infants to fixate on one location on the screen in response
18 to the presence of a sound and used visual rewards to reinforce the looking behavior. Of 6
19 infants who were successfully calibrated, only 2 passed the training criterion, and neither of
20 these completed the following test phase at above-chance level.

21
22 Like infants, ASD children with impaired receptive language may have limited understanding of
23 verbal instructions and have trouble providing verbal or manual responses. However, these
24 ASD children are of course not infants – they are children who probably have experience with
25 interacting with electronic devices; who can potentially concentrate for longer (given the right
26 task); who have greater control of their eye-movements, and who usually have some
27 understanding of cause and effect regarding their impact on the environment. As such, we
28 aimed to learn from reports with infant participants and adapt the methods to achieve even
29 greater success in developing a gaze contingent paradigm for minimally verbal ASD children.

30 31 **Aims**

32
33 In the current study, (1) we develop a gaze contingent test for auditory perception that could be
34 completed by typical children and those with severe impairments in language and intellectual
35 impairment ($IQ < 70$). Our paradigm requires them to discriminate between two sounds. At the
36 easiest level, the two sounds were the most extreme examples; at the hardest level, the two
37 sounds were the most similar perceptually. This allows us to record a perceptual threshold. (2)
38 With such a test we can compare auditory difference thresholds for stimuli including pitch and
39 rhythm (using both human voice and synthetic sounds). (3) Lastly, by open sourcing our Matlab
40 test code it can be used and improved by the research community.

41 42 **Method**

43 44 **Participants**

45
46 A group of 16 typically developing children (TD; age range: 4.4 – 9.9 years; mean age: 8.4
47 years) and 9 ASD children (age range: 5.5 – 9.9 years; mean age: 8.2 years) from the region
48 near Seville, Spain were tested. ASD children were recruited through 'Autismo Sevilla', an
49 organization for autistic individuals, and 'Asociación SETA' a center for attention and early
50 intervention. Diagnostic reports were obtained via the parents from clinical psychologists or
51 psychiatrists. If not previously conducted, the Autism Diagnostic Observation Schedule-Generic

(ADOS-G, Lord et al. 2000) was completed by the research team. Minimally verbal participants completed module 1, the remainder completed module 2. The ADOS was not completed with one participant due to anxiety and noncompliance. Because he already had a clinical diagnosis, and observations by the research team who are experienced with autism strongly indicated ASD, it was decided to terminate the ADOS assessment. The Spanish version of the Autism Quotient (AQ; Auyeung et al., 2007) was completed for all participants, although final scores for one ASD participant and 6 TD participants were not available due to missing items or questionnaire not being returned. If a participant scores over 76, an assessment of ASD is recommended. However, it is not a diagnostic tool and the four ASD participants we had who scored below this cut-off (see Table 1) were still analyzed as within the spectrum as they had scores within the autism range on the ADOS. Parents of all children reported that their child had no hearing difficulties, and completed the Spanish version of The Child Sensory Profile 2 (Dunn, 2014), providing corroborative information on participants' sensory experiences.

To measure receptive language, the Spanish version of the Test for Reception of Grammar (TROG; Bishop, 2003; Spanish version: 'Test de Comprensión de Estructuras Gramaticales' (CEG), Mendoza 2005, Mendoza et al. 2005) was attempted with all participants, although it was not possible to complete the test with one ASD participant. To measure nonverbal cognitive ability, the Kaufman Brief Intelligence Test matrices (KBIT; Kaufman and Kaufman 2004) subtest was completed for all participants. This task was considered one of the most effective measures of nonverbal IQ in this population because participants are not required to give a verbal response.

Of the ASD participants, 3 were minimally verbal (they had fewer than five productive words to communicate and receptive language scores in the impaired range), 4 participants had five or more productive words to communicate and scored in the impaired range on the receptive language task, and 2 had fluent phrase speech and scored in the normal range on the receptive language task. Details of participant demographics and cognitive measures are provided in Table 1.

[Insert Table 1]

Table 1. Individual Behavioral measures and Auditory Threshold Results. Participant diagnosis, age, IQ scores, autistic traits, sensory profile and behavioral results for our auditory threshold game are presented. Blank entries indicate the participant was not tested. ** Two typically developing participants completed an older version of the task in which training phases 1A & 1B had a pass criterion of 3 out of 3 correct trials. The testing phase and all parameters and stimuli were identical.

Stimuli and set-up

Our source code and stimuli are located at:
<https://github.com/VisionResearchBlog/gaze-audio-threshold>

Participants were seated ~60cm from the eye tracking setup. For ease of calibration, we used a Tobii TX300 tracker that consists of an eye tracker camera base with a 23" monitor on top (the screen subtends roughly 45°x25° of visual angle at that viewing distance). Note, our setup is portable to other trackers and was also tested on a Tobii X2-60, which is smaller and cheaper, but runs at a lower frame rate (but still sufficient for our gaze contingent testing). Conversion to use a different brand of tracker is possible but would require sufficient programming skill as the current software only supports trackers using the Tobii Pro software development kit. Stimulus

1 presentation and data recording were controlled by a PC running Matlab R2015a (Mathworks,
2 Natick, MA, USA).

3
4 The testing environment was in a quiet lab space (without sound proofing) with overhead
5 lighting. 'Attention getting' calibration targets designed for children (included with Tobii Studio
6 software) were used to calibrate the tracker. The animations subtended $1.2^{\circ} \times 1.2^{\circ}$ and were
7 placed in a randomized sequence in five locations onscreen (upper left, bottom left, center, top
8 right, and bottom right) to gather calibration data. The brief 1.5s animations of items included
9 drawings of toys including a bus, caterpillar, dog, chick, cat, lobster, and rattle, all paired with
10 brief semantically unrelated sounds. For example, the toy bus shrinks and grows while one
11 hears the sound of an orchestral fanfare, and the chick shakes right and left with the sound of a
12 two-note piano ostinato.

13
14 Auditory stimuli were in 4 sets of 10 fixed examples that varied within one perceptual dimension.
15 The four sets included pitch (synthetic and human voice) and rhythm (synthetic and human
16 voice). The pitch stimuli were single sounds of 2s duration at a frequency of: 415Hz (G[#]);
17 428Hz; 440Hz; 453Hz; 466Hz; 473Hz; 480Hz; 488Hz; 492Hz; 494Hz (B). The rhythm stimuli
18 were a series of beats lasting 6-7s, at a fixed frequency of 330Hz in the synthetic set and 196Hz
19 in the human set. The beat patterns gradually and systematically increased in complexity. At
20 each progression, one bar out of three would change slightly. The first stimulus was 3 bars of 4
21 quarter-notes. The second stimulus was 2 bars of 4 quarter-notes, and 1 bar consisting of 1
22 quarter-note / 2 eighth-notes / 1 half-note. The tenth stimulus was 3 identical bars each
23 consisting of 3 eighth-notes / 2 sixteenth-notes / 1 half-note. The rhythm combinations are
24 visually represented as a musical score in the file set which is free to download via the link
25 above ('Rhythm Beats.pdf'). The synthetic pitch stimuli were generated using Praat
26 (www.fon.hum.uva.nl/praat/, Boersma & Weenink, 2018; Boersma (2001)) and Adobe Audition
27 7.0 (Adobe, San Jose, CA, USA, 2014). The synthetic rhythm stimuli were generated in
28 Noteflight (www.noteflight.com, Noteflight, LLC, Boston, MA, USA, 2007), an online music
29 notation software, and mimicked the sound of a piano. The human versions were of an adult
30 female mimicking the synthetic versions and saying the syllable 'ba'. They were recorded in a
31 soundproof lab edited afterwards using Praat so that the final versions were as close as
32 possible to their synthetic counterparts whilst maintaining a human quality. All stimuli were
33 presented at 60dB measured using a sound level meter (Velleman, DVM1326). All sounds were
34 played using a single speaker (Sony, SRS-GU10iP) positioned in the center directly below the
35 monitor.

36
37 Visual stimuli were hand drawn characters – a dinosaur, alien and a snail ($2.8^{\circ} \times 2.8^{\circ}$ square).
38 For each character, 12 different silent 'reward animations' of ~1s were generated using Adobe
39 Photoshop, in which the character completed a 'positive' and entertaining action, e.g. dancing,
40 smiling, waving. Two of the animations were designed to be particularly rewarding and
41 encouraging, e.g. silent animations of the character cheering with multicolored fireworks in the
42 background.

43
44 All individual auditory and visual stimuli can be downloaded from the source code link provided
45 above. Note, an additional set of sounds varying in volume were generated to test absolute
46 thresholds, but were not tested as the experimental sessions were becoming too long (the
47 duration of sessions varied from 20 – 45 minutes). However, these sounds are also available in
48 the stimulus set provided in the link.

49
50 Participants completed two versions of the task (one pitch, one rhythm) in a session, and the
51 order was counterbalanced across participants. The experimenter was always present in the

room and she offered support and guidance during calibration, and to help the participant grasp the concept of the point-of-gaze fixation with their eye gaze (see below). Once the experiment started she was silent, only interrupting if the quality of the track was poor and recalibration was necessary. ASD participants completed the experiment with a parent also present. This was necessary to reduce anxiety and ensure the participant remained seated and attentive to the task. The parent was told to not point at any images on the screen, and to only help the participant maintain attention towards the screen. For control participants it was not necessary to have a parent present.

Trial protocol

Our experimental design went through many revisions (see discussion on design choices). The final version consisted of an initial eye tracker setup followed by four phases that gradually introduced the participant to gaze-based interaction in the context of a 3 up – 1 down staircase. First, participants were seated in front of the eye tracking and monitor setup and calibrated. After calibration, the software would display the mean and standard deviation of the calibration data for each calibration point and overall. We sought to achieve error below 2-3 degrees. Throughout the experiment a small orange square ($0.25^{\circ} \times 0.25^{\circ}$) was rendered in real-time at the current point-of-gaze; the participant was encouraged to recognize that this represented their point-of-gaze by making them follow a finger of the experimenter moving around the screen. All participants grasped the concept of manipulating the orange square with their eye-gaze. This also meant that the experimenter (who silently observed behind the participant) had direct feedback on the quality of the track and was able to pause the program and recalibrate if needed.

The initial stimulus presentation screen used a grey background with three zones. This consisted of a central fixation zone displaying one of the characters and two rectangles ($7.1^{\circ} \times 5.7^{\circ}$), a blue one on the left and a red one on the right, both with a central white question mark ($1.9^{\circ} \times 1.9^{\circ}$). Real-time area of interest detection used a generous 4° boundary around central fixation and the 2 rectangles.

Figure 1 shows how the trials progressed, with details of the variations from training through to main experiment protocol described in the section below. The trials always started with the presentation of a character in the central fixation zone on which the participant was required to fixate for 300ms. If the program detected the participant had not fixated the central zone after 2s had elapsed, the character would move left and right ($\pm 2^{\circ}$) as an attention getting stimulus. Once the participant fixated, the auditory stimulus played and the two rectangles appeared.

Throughout the experiment, the inter-trial interval – defined as the time between the end of the final visual presentation of one trial (see Figure 1A & B) and the appearance of the central fixation image for the next trial – was a number randomly selected between 0.5s and 3s using Matlab's *rand()* function.

[Insert Figure 1A & B – Screen capture of trial progression]

Figure 1: (A) Starting from the top, showing the progression in time of the onscreen sequence for Phase 1A & 1B training trials. (B) The time progression for Phase 2 training trials and main experimental testing trials. Orange square represents the gaze cursor that participants learnt to control by moving their eye-gaze.

1 The crux of the game was that participants associated one stimulus with the left-hand side, and
2 another stimulus with this right-hand side (this stimulus-response pairing was kept consistent
3 per participant, but randomized across).

4 ***Training Phases and Main Experiment Protocol***

6
7 The auditory threshold experiment consisted of two training phases that the participant had to
8 pass in order to move onto the full staircase threshold test.

9
10 *Training Phase 1A:* This phase taught the pairing of stimulus and location, e.g. the highest tone
11 of the 10 pitch stimuli was paired with the left-hand side. Figure 1A shows the progression of
12 Training Phase 1. Once the subject had fixated the central fixation and the auditory stimulus had
13 played, the central fixation character was animated to listen to the left and right by holding their
14 hands to their ears and looking each way. Next, the central fixation disappeared and the
15 'correct' rectangle flashed twice to draw attention. If the participant then looked at the correct
16 rectangle for the minimum duration (333ms), the reward animation was triggered. If the
17 participant looked at the wrong rectangle for the minimum duration (333ms), the rectangle
18 turned black with a cross over it. If the participant did not look at either rectangle within 6s, the
19 trial timed out, the screen dimmed for 1.5s and the rectangles turned blank grey. All stimuli then
20 disappeared, and the next trial started once the random inter-trial interval (0.5-3s) elapsed.

21
22 In this phase, the participant was required to look to the correct location in the allotted time on 2
23 of 2 trials. If they did not meet this criterion, they were presented with another set of 2 trials and
24 this evaluation was repeated as necessary.

25
26 *Training Phase 1B:* In this phase the set-up was the converse of Phase 1A stimulus pairing, e.g.
27 the lowest tone of the set of 10 pitch stimuli was paired with the right-hand side.

28
29 Training phases 1A & 1B were then repeated once apiece.

30
31 *Training Phase 2:* Figure 1B demonstrates the sequence of events. This phase continued to
32 present only the two extreme stimuli (e.g. highest vs. lowest pitch), but the 'listening' animation
33 was not used, nor did the rectangles flash after the sound was played. The participant needed
34 to use the sound & location mapping they learned in Phase 1 to pass trials. After hearing the
35 auditory stimuli, the two rectangles were presented and participants were expected to fixate one
36 or the other. As before, if they fixated the correct rectangle for the minimum duration (333ms)
37 they viewed a reward animation; if they fixated the wrong rectangle they saw a cross; if they
38 fixated neither rectangle within 6s, the screen went grey for 500ms and the trial ended and
39 proceeded to the next trial. Six trials (3 of each auditory stimulus) were randomly presented and
40 they needed to fixate the correct rectangle on 5 of 6 trials. If they did not meet this criterion, they
41 were presented with another set of 6 and this evaluation was repeated as necessary.

42
43 The total minimum number of training trials to complete before progressing to the full
44 experiment was 14. No maximum trial number was set, but if the participant clearly was not
45 grasping the concept, the experiment was terminated.

46
47 *Testing Phase:* Here the sequence of visual events was identical to Training Phase 2 (Figure
48 1B), but now the three-up/one-down staircase procedure was introduced to move through the
49 auditory levels. As the participant progressed through the trials, when they went 'up' this
50 triggered one of the stimuli to increment one step closer in perceptual similarity. During the
51 entire experiment, one stimulus would shift whereas the other would remain the same acting as

a static reference. There were three possible ways for the testing phase to complete. 1) Repeated successful discrimination at the highest level of difficulty four trials in a row. This means their auditory difference threshold is at the highest level. 2) Repeated unsuccessful discrimination at the lowest level of difficulty six trials in a row. This means their auditory difference threshold is at the lowest level or there is some other confound preventing their success. 3) If the subject achieved 11 reversals, that is going up and then down in difficulty, or down then up, the program would come to a halt. We assume that the highest stimulus level at which the reversals occur represents the participant's difference threshold.

At the easiest level the two sounds were the most extreme examples; at the hardest level the two sounds were the most perceptually similar. Therefore, in the pitch experiments, participants distinguished between 415Hz and 494Hz in level 1, and between 492Hz and 494Hz in level 9. In the rhythm experiment, level 1 required participants to distinguish between 3 bars of equal beats and 3 bars of complex beat combinations. In level 9, both stimuli contained 3 bars of complex short and long beats and they differed only at one point where two eighth notes replaced a single quarter note.

On 1 in 10 successful trials, in addition to the character animation, participants would see an additional brief animation of gold going into a pot, jelly beans into a jar, or a star emerging from a magical box. This was intended to provide variety and entertainment for the participant to keep them engaged to perform correctly. It was decided not to use sounds as a reward to avoid interfering with learning the auditory stimuli. If the participant was very disinterested, we set up the experimental program so that a key was mapped to pause the experiment and play a silent video (in our case an episode from a child's animation program).

Experiments were counterbalanced within and between subjects in terms of side of static reference stimuli (i.e. left or right) and pairing of auditory and visual stimuli (e.g. dinosaur with pitch). Selection of experiment parameters and counterbalancing were entered as needed by the experimenter into the Matlab code before each trial block.

Results

Several adults completed pilot tests and demonstrated that they could perform the audition game at or near the highest level of difficulty (level 9).

All TD and ASD children were reported to have normal hearing and no history of hearing impairment, although none underwent formal testing. Table 1 provides details of individual participants' behavioral measures and their performance for the threshold experiment, including the highest level reached and the number of trials completed. As shown in the table some children were not able to exit training for some stimulus sets. Testing was not attempted for every version of the task in every child. This was due to time restraints, or because the participant was clearly unable to grasp the concept of the experiment.

All participants were able to complete several trials by manipulating the orange square on the screen with their eye-gaze to select one of the two boxes.

TD participants

In summary, all 16 TD children were able to pass training for at least one of the stimulus sets. The number of trials taken to achieve this ranged from the minimum of 14 (taking between 1-2

minutes to complete) to 110 (taking around 10 minutes to complete). Note that prolonged times during training most likely tired our participants and in extreme cases like the latter of 110 trials, it may be that training success was due to luck and the participant was not clear on the rules of the game.

TD threshold scores vary from the lowest level to the highest and are quite heterogeneous. An overall mean score was calculated for each participant averaging across all attempted versions, and allocating a '0' score when training criteria was not met. Mean score across all TD participants was 3.6 (s.d. 2.8). Despite our small participant pool, which is underpowered for most statistics, we analyzed correlations. Overall mean scores were tested for a correlation with age, receptive language normed score, and nonverbal IQ normed score. Using adjusted alpha levels of 0.017, there were no significant associations between mean threshold score and age ($r(16) = 0.14$, $p = 0.6$); receptive language ($r(16) = 0.48$, $p = 0.06$); or nonverbal-IQ ($r(16) = 0.34$, $p = 0.2$). An example of a TD participant's performance on a synthetic rhythm experiment test is presented in the left-hand panel of Figure 2.

Next, mean scores were calculated for 'pitch', 'rhythm', 'human' and 'synthetic', by taking the average highest-level when two experiments in the category were completed, or using the available highest-level if only one experiment was completed (and including '0' when the training phase was not passed). To determine if the versions of the experiment were of comparable difficulty, a 2x2 ANOVA was conducted with stimulus types as levels: 'pitch / rhythm' and 'synthetic / human'. The interaction was non-significant ($F(1,15) = 1.96$, $p = 0.18$, $\eta^2 = 0.12$), as were the main effects of pitch / rhythm ($F(1,15) = 0.01$, $p = 1.00$, $\eta^2 = 0.01$) and synthetic / human ($F(1,15) = 0.11$, $p = 0.75$, $\eta^2 = 0.01$).

ASD participants

Only 5 of the 9 ASD children passed training for at least one of the stimulus sets. In the ASD participant group the mean threshold score was 1.8 (s.d. 3.0). Unfortunately, too few ASD children passed the training trials to conduct any meaningful analyses on the results.

In Table 2, we expand briefly on our experiences of testing each of these individuals.

[Insert Table 2]

Table 2. ASD participants and common problems encountered during their test runs. Problems are coded from 0 to 5, with: 0 = Anxiety / noncompliance; 1 = Did not understand the game; 2 = Problems with generalizing; 3 = Quantity of testing; 4 = Participant lost interest; 5 = Problems with memory

[Insert Figure 2 showing performance and threshold estimate]

Figure 2: Two examples of trial-by-trial performance on the test phase. In the 3-up, 1-down regime, each step increase is the result of 3 correct choices, and each decrement is due to 1 wrong choice. In the left panel, the performance of TD participant ('C01') is presented on the synthetic rhythm task. This participant completed 78 training trials before progressing to the main experiment; he then completed 44 test trials, of which 32 were correct. His threshold is estimated at level 5. In the right panel an ASD participant's performance ('A04') on the human pitch task is presented. This participant completed 16 training trials before progressing to the main experiment; he then completed 49 test trials, of which 42 were correct. His threshold is estimated at level 8 (i.e. discrimination between 488Hz and 494Hz).

Discussion

We have demonstrated a proof of concept gaze contingent method to evaluate auditory thresholds in the typical child population. Unfortunately, performance was mixed in both groups and often unsuccessful with the atypical children we tested. We were forced to question whether the participants had auditory impairments that prevented them from perceiving auditory differences, or if the game was too complex or confusing for our participants. Furthermore, it is unclear whether the varied performance of the control group represented real differences in auditory perception abilities, or whether the introduced variance was a product of the task. To address these concerns there is a need for corroborating information on auditory thresholds from another methodology. Traditional VRA methods with control participants would be a valuable first step, but as this is a problematic methodology in autistic participants, our group is pursuing the use of auditory evoked potentials as a baseline to compare against our method (Ruiz Martinez, Wilson, Yau, Saldaña et al. under review).

Despite apparent limitations of the paradigm, we document some notable successes. First, all participants could be calibrated. This is no small achievement as some participants were had severe intellectual impairments with significant behavioral, cognitive and motor disabilities. The infant calibration method (code for which is provided in the GitHub link) and the high-resolution eye-tracking system were both instrumental to this success. Second, the use of a visible gaze cursor that participants learnt to control was extremely effective, and all participants – regardless of age and level of IQ – successfully did this without being verbally instructed. As such, we have demonstrated that the use of eye-tracking methodology does not need to be passive, but that minimally verbal children can interact with experiments and effectively provide responses. This is a significant contribution to methodology and opens avenues for conducting much needed research in this challenging and neglected population. Third, as a result of pilot testing and several modifications, the final versions of the stimuli used in the different ‘games’ were of comparable difficulty for typical children. The different versions could therefore be used to compare perceptive ability of pitch versus rhythm information, as well as human versus non-human sounds. Adaptations for testing other auditory qualities (e.g. volume and duration, stimuli available in GitHub link) could easily be incorporated.

Reflections and future directions

If we assume that the auditory impairments of our participants were not so severe that they could not perceive the differences in even the most extreme stimuli, then we must accept that the difficulties progressing through the experimental ‘game’ were due to limitations in the protocol. Given the sparsity of our data it is difficult to unravel issues due to stimulus choice, game design, and idiosyncrasies of each participant. Auditory difference thresholds may be attainable with the right combination of stimuli, design and higher function participants, but more coarse discrimination may be possible in other groups. Alternatively, if the task proves too difficult to accurately measure thresholds, the experimental set-up may prove more useful to compare perception of different stimulus types in a within-subject design and between groups.

Nevertheless, we remain hopeful that with careful design our setup could work for auditory differences thresholds. Following the summary of specific challenges faced by each participant (Table 2), we discuss these insights on potential reasons for lack of success and provide ideas for further modifications.

1 1. *Participants did not understand the game.* The object of the game is for the participant to
2 learn to associate one sound category with one location on the screen, and another sound
3 category with another location on the screen. This concept may have simply been too
4 complicated for some of our participants, or we may have asked too much in expecting them to
5 master the association so quickly. Additionally, our choice of reinforcing stimuli may not have
6 been well tuned for our participants thus inhibiting them from learning the correct associations.

7
8 2. *Problems with generalizing.* The test assumes that the child will generalize the rules learned
9 in training phases when confronted with a novel stimulus that is similar to one of the categories
10 they learned. It may be the case that this is confusing, as they do not encounter the
11 generalization stimuli in the training phases and they were unable to grasp the connection
12 between the novel stimulus and the original. Given that individuals with ASD are thought to have
13 difficulties generalizing (Klinger & Dawson, 2001), this may adversely affect their performance
14 (e.g. perhaps the case with participant A05). Perhaps focusing on the training of one 'anchor'
15 stimuli, and later introducing the alternative stimuli, would alleviate this problem. This would
16 create more of a 'same / different' testing paradigm.

17
18 3. *Quantity of testing.* A key aim was to test multiple dimensions of auditory stimuli using the
19 same protocol. Unfortunately, this may have introduced problems as participants were expected
20 to learn first one set of stimuli (e.g. pitch) then another (e.g. rhythm). A more effective learning
21 procedure could have focused on one dimension and perhaps completed a training session on
22 one day to familiarize the participant with the concept of the game and learn a visual-auditory
23 association, then returned to complete the experimental task on the following day. Of course,
24 introducing methods that require participants and their families to attend multiple sessions at
25 fixed times can be problematic, but perhaps necessary to achieve success.

26
27 4. *Participants lost interest.* Despite our best efforts to create an interesting 'game', the
28 repetitive experimental nature may still have resulted in loss of interest. Nevertheless, during
29 development of the software we did have several ideas to increase the level of interest – some
30 of which were adopted, others were rejected (see 'discussion on design choices' below). To
31 improve this perhaps some game elements could be emphasized, e.g. keeping score, or more
32 attractive graphics and animations could be created.

33
34 5. *Problems with memory.* If our participants did not remember the rules or became confused as
35 to when they applied this would result in poor performance. Additionally, the game requires a
36 mental comparison between the current stimulus and the anchor stimulus before a choice can
37 be made. At each step interval, odds are the participant will hear the anchor stimulus, but this
38 was not guaranteed as stimulus selection was randomized. This presentation style may have
39 led to scenarios where the participant had forgotten the anchor and could not correctly compare.
40 In our opinion, this accounts for why some of the typical participants were able to reach high
41 levels on some versions of the game, but could not pass training criteria on others: the problem
42 was not that they could not discriminate between the extreme examples of stimuli, but more
43 likely that they made mistakes and then became confused and were unable to re-establish
44 which auditory stimuli was associated with which rectangle. Again – a solution may be to focus
45 on the training of the anchor stimuli, and only once this is well established to introduce the
46 alternative stimuli and test discrimination against the anchor.

47
48 As is clear from Table 2, the elements of the task that the ASD children found challenging were
49 extremely varied: some were anxious or noncompliant, others were not; some were motivated
50 by the reinforcers and enjoyed the 'game', but others did not; some easily grasped the audio-
51 visual association, others never grasped it. Moreover, it was not necessarily the children with

1 fluent speech or those with average IQ that were successful. This extreme variability in
2 performance supports the use of individualized approaches advocated by some previous
3 researchers (e.g. Plesa Skwerer et al. 2016; Kasari et al. 2013). Perhaps combining the core
4 elements of the testing paradigm, with individualized reinforcers and/or cues to encourage rapid
5 grasping of the concept, would yield better task success and more reliable indication of auditory
6 discrimination ability.

9 ***Discussion on design choices***

11 During the process of developing the game, several edits were made to stimuli and protocol.
12 These followed initial pilot testing with adults (prompting revisions in stimuli (points 1 and 2)),
13 and subsequent tests with typical children (leading to revisions in training and test protocol
14 (points 3 and 4)). The key design choices are outlined here so that other researchers may
15 benefit from lessons that were learnt.

17 1. *Pitch stimuli*. Originally, the pitch stimuli ranged from 330Hz ('E') to 587Hz ('D'), with a
18 semitone between each level. The large differences between levels were problematic because
19 participants found it difficult to generalize across incrementing stimuli during the staircase
20 phase. Also, the hardest level – intended to capture threshold discrimination – would have been
21 clearly different to most people of normal hearing. Therefore, stimuli were modified so that
22 differences between levels were much smaller and at the most difficult level the step was 2Hz –
23 barely perceptible to most individuals of normal hearing.

25 2. *Rhythm stimuli*. Originally rhythm stimuli were 3.5s long, to be more in line with pitch stimuli.
26 However, it proved to be much more difficult than the pitch stimuli because of the rapid
27 presentation of beats. In order to make the versions of similar difficulty level, the rhythms were
28 slowed down, although this was at the cost of making the experiment much longer.

30 3. *Training phases*. The training phases underwent several modifications, with the constant
31 challenge of training participants as efficiently as possible – any method that took too long
32 would risk boring the participant; any method that had too many variables in it would risk
33 confusing the participant.

35 First, we attempted showing only one rectangle in the earliest phases, but pilot tests indicated
36 that participants were immediately confused when two rectangles were shown, probably
37 because they had not realized the importance of the auditory stimuli until that point. Therefore,
38 both stimuli were always present, but in Training Phase 1 the correct side flashed to draw
39 attention to it.

41 Second, we created a training scenario where after the audio stimulus was played the central
42 fixation character appeared to listen, by holding up hands to ears, and would then automatically
43 slide over to the correct side. After successful trials the character would no longer automatically
44 slide over to the correct location but was instead controlled by the child's gaze. However, here
45 the child often failed to learn the gaze contingency and would stare at the central fixation waiting
46 for the character to move. Therefore, we included the animation of the character listening to
47 each side to help the participant understand that they must attend to the sound, but did not
48 move the character automatically. It is possible that further non-verbal instruction may have
49 helped, while we did not implement this, it would be possible to include a graphic cue that
50 indicates what the child should pay attention to, for instance when testing pitch an arrow
51 pointing up or down might be useful.

1
2 Third, we piloted versions with far more trials in Training Phase 1 (blocks of the most extreme
3 stimuli) but found that participants became bored, and also forgot the first stimuli when listening
4 to the second. In order to balance the presentations whilst providing enough repetitions to learn
5 the distinct stimuli, 2 sets of 2 examples on each side was decided on. Fourth, we tried using
6 much longer inter-trial intervals so that learning of stimuli could be assimilated, but participants
7 quickly lost interest.
8

9 4. *Test phase.* We considered having the child's progress be represented onscreen, either by
10 having the extra reward animations with gold, jelly beans, or stars have a marker to stay
11 onscreen to show how many awards have been collected; this was not done to avoid visual
12 clutter. We also considered having the onscreen stimuli start at the bottom of the screen and
13 move upwards whenever the child progressed a step up in the level of difficulty, but we
14 reasoned this may be confusing to the child. Using sound stimuli to encourage participants or
15 re-enforce success or failure of a trial was also considered but rejected because it may interfere
16 with learning the target sounds stimuli.
17

18 **Conclusion**

19

20 There are many aspects to consider when turning a psychophysical test into a game that
21 children of a variety of abilities have the capacity to enjoy and easily learn and interact with.
22 Despite apparent limitations in our approach, we have made significant progress in this regard
23 and document our efforts so that the field may continue to move forwards. The code we
24 designed, and the well-controlled auditory and visual stimuli, are now freely available to other
25 researchers to use and adapt.
26
27

Acknowledgements

CEW was supported by a Talentia Postdoc launched by the Andalusian Knowledge Agency, co-funded by the European Union's Seventh Framework Program, Marie Skłodowska-Curie actions (COFUND – Grant Agreement nº 267226) and the Ministry of Economy, Innovation, Science and Employment of the Junta de Andalucía, Spain. BTS was supported by the Marie Skłodowska-Curie actions LanPercept Innovative Training Network (FP7-PEOPLE-2012-ITN, Project ID: 316748).

We thank Joel Talcott and Santiago Benjumea for helpful discussions on method design. Also to Marina Barrio for her assistance in producing the sound stimuli.

Finally, we are extremely grateful to all the participants and their families that participated in this study.

Bibliography

- Auyeung, B., Baron-Cohen, S., Wheelwright, S., Allison, C. 2007. The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of autism and developmental disorders*, (38), pp.1230–1240.
- Bouton, S., Bertoncini, J., Serniclaes, W., & Colé, P. (2011). Reading and reading-related skills in children using cochlear implants: Prospects for the influence of cued speech. *Journal of deaf studies and deaf education*, 16(4), 458-473.
- Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.
- Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.27, retrieved 17 March 2017 from <http://www.praat.org/>
- Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. M. (2003). Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *Journal of cognitive neuroscience*, 15(2), 226-235.
- Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., ... & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, 48(9), 2465-2475.
- Brueggeman, Paul, Aud, CCC-A. 10 Tips For Testing Hearing in Children With Autism. The ASHA Leader, January 2012, Vol. 17, 5-7.
- Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, 111(40), 14559-14564.
- Downs, D., Schmidt, B., & Stephens, T. J. (2005, November). Auditory behaviors of children and adolescents with pervasive developmental disorders. In *Seminars in Hearing* (Vol. 26, No. 04, pp. 226-240). Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
- Dievendorf, Allen O. and Gravel, Judith S. (1996) Behavioral Observation and Visual Reinforcement Audiometry. In S.E. Gerber (ed.), *The Handbook of Pediatric Audiology*, chapter 4, 55-83. Gallaudet University Press.
- Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, 95(2), 1053-1064.
- Egelhoff, K., Whitelaw, G., & Rabidoux, P. (2005, November). What audiologists need to know about autism spectrum disorders. In *Seminars in Hearing* (Vol. 26, No. 04, pp. 202-209). Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
- Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS comput biol*, 5(3), e1000302.
- Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers in human neuroscience*, 8, 777.

- 1
- 2 Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J.
- 3 (2015). Musical rhythm discrimination explains individual differences in grammar skills in
- 4 children. *Developmental Science*, 18(4), 635-644.
- 5
- 6 Heaton, P., Williams, K., Cummins, O., & Happé, F. (2008). Autism and pitch processing splinter
- 7 skills: a group and subgroup analysis. *Autism*, 12(2), 203-219.
- 8
- 9 Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time
- 10 perception and developmental dyslexia: perception of musical meter predicts reading and
- 11 phonology. *Cortex*, 47(6), 674-689.
- 12
- 13 Jones, C. R., Happé, F., Baird, G., Simonoff, E., Marsden, A. J., Tregay, J., ... & Charman, T.
- 14 (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum
- 15 disorders. *Neuropsychologia*, 47(13), 2850-2858.
- 16
- 17 Kaga K, Tanaka Y. Auditory Brainstem Response and Behavioral Audiometry. *Developmental*
Correlates Arch Otolaryngol. 1980;106(9):564-566
- 18
- 19 Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal
- 20 school-aged child with autism spectrum disorder. *Autism Research*, 6(6), 479-493.
- 21
- 22 Kaufman, A.S. & Kaufman, N.L. (2004). KBIT: Kaufman Brief Intelligence Test (KBIT
- 23 SpanishVersion). Madrid: TEA Editions.
- 24
- 25 Klintwall, L., Holm, A., Eriksson, M., Carlsson, L. H., Olsson, M. B., Hedvall, Å., ... & Fernell, E.
- 26 (2011). Sensory abnormalities in autism: a brief report. *Research in developmental*
disabilities, 32(2), 795-800.
- 27
- 28 Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and
- 29 linguistic processing of speech in preschool children with autism: behavioral and
- 30 electrophysiological measures. *Developmental science*, 8(1), F1-F12.
- 31
- 32 Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M.
- 33 (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social
- 34 and communication deficits associated with the spectrum of autism. *Journal of autism and*
developmental disorders, 30(3), 205-223.
- 35
- 36
- 37 Mendoza, E. (2005). CEG: Test de Comprensión de Estructuras Gramaticales. Madrid: TEA
- 38 Editions. Madrid: TEA Editions.
- 39
- 40 Mendoza, E., Carballo, G., Muñoz, J., & Fresneda, M. D. (2005). Evaluación de la comprensión
- 41 gramatical: un estudio translingüístico. *Revista de Logopedia, Foniatría y Audiología*, 25(1), 2-18.
- 42
- 43 Miyazaki, M., Takahashi, H., Rolf, M., Okada, H., & Omori, T. (2014). The image-scratch
- 44 paradigm: a new paradigm for evaluating infants' motivated gaze control. *Scientific reports*, 4.
- 45
- 46 Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language
- 47 learning. *Proceedings of the National Academy of Sciences*, 109(39), 15953-15958.
- 48

- Musiek, FE & Baran, JA (2007). *The Auditory system*. Boston, MA: Pearson Education, Inc.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12), 2544-2590.
- Paulraj, M. P., Subramaniam, K., Yaccob, S. B., Adom, A. H. B., & Hema, C. R. (2015). Auditory evoked potential response and hearing loss: a review. *The open biomedical engineering journal*, 9(1).
- Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, 20(5), 591-604.
- Renshaw, J.J. and Diefendorf, A. O. (1998) Adapting the Test Battery for the Child with Special Needs. In F. H. Bess (ed.), *Children with Hearing Impairment: Contemporary Trends*, chapter 7, 83-104, Vanderbilt Bill Wilderson Center Press.
- Rosenhall, U., Nordin, V., Sandström, M., Ahlsen, G., & Gillberg, C. (1999). Autism and hearing loss. *Journal of autism and developmental disorders*, 29(5), 349-357.
- Sabo, D. L. (1999). The audiologic assessment of the young pediatric patient: the clinic. *Trends in amplification*, 4(2), 51-60.
- Samson, F., Mottron, L., Jemel, B., Belin, P., & Ciocca, V. (2006). Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks?. *Journal of autism and developmental disorders*, 36(1), 65-76.
- Schall, U. (2016). Is it time to move mismatch negativity into the clinic?. *Biological psychology*, 116, 41-46.
- Schwarz, I. C., Nazem, A., Olsson, S., Marklund, E., & Uhlén, I. (2014). Towards a contingent anticipatory infant hearing test using eye-tracking. In *FONETIK 2014, the XXVIIth Swedish Phonetics Conference* (pp. 35-40). Department of Linguistics, Stockholm University.
- Schwartz, S., Shinn-Cunningham, B., & Tager-Flusberg, H. (2018). Meta-analysis and systematic review of the literature characterizing auditory mismatch negativity in individuals with autism. *Neuroscience & Biobehavioral Reviews*, 87, 106-117.
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural brain research*, 291, 244-252.
- Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism*, 18(2), 137-147.
- Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of autism and developmental disorders*, 44(12), 3161-3167.
- Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Research*, 6(6), 468-478.

1
2 Thompson, G. Wilson, R.W. and Moore, J.M. (1979). Application of visual reinforcement
3 audiometry (VRA) to low IQ children. *Journal of Speech and Hearing Disorders*, 44(2), 80-90.

4 Wang Q, Bolhuis J, Rothkopf CA, Kolling T, Knopf M, Triesch J: 'Infants in control: rapid
5 anticipation of action outcomes in a gaze-contingent paradigm', *PLoS ONE*, 2012
6

7 Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy.
8 *Current biology*, 21(18), 1543-1547.

9 Watson, L. R., Roberts, J. E., Baranek, G. T., Mandulak, K. C., & Dalton, J. C. (2012).
10 Behavioral and physiological responses to child-directed speech of children with autism
11 spectrum disorders or typical development. *Journal of autism and developmental*
12 *disorders*, 42(8), 1616-1629.
13

14 Wilson, W.R., Folsom, R.C., and Widen, J.E. (1983) Hearing impairment in Down's syndrome
15 children. In G. Mencher and S. Gerger (eds), *The multiply handicapped hearing impaired child*.
16 New York: Grune and Stratton.

17 Yau, S. H., McArthur, G., Badcock, N. A., & Brock, J. (2015). Case study: auditory brain
18 responses in a minimally verbal child with autism and cerebral palsy. *Frontiers in*
19 *neuroscience*, 9.
20
21